Characterization of novel PAH dioxygenases from the bacterial metagenomic DNA of a contaminated soil.

Angelina Chemerys, Eric Pelletier, Corinne Cruaud, Florence Martin, Fabien Violet, Yves Jouanneau

Characterization of novel PAH dioxygenases from the bacterial metagenomic DNA of a contaminated soil

Angelina Chemerys[1,2,3], Eric Pelletier[4,5,6], Corinne Cruaud[4,5,6], Florence Martin[1,2,3], Fabien Violet[1,2,3] and Yves Jouanneau[1,2,3]

[1]CEA, DSV, Laboratoire de Chimie et Biologie des Métaux, F-38054 Grenoble, France

[2]CNRS UMR 5249, 38054, Grenoble, France

[3]Univ. Grenoble Alpes, 38042, Grenoble, France

[4] CEA, DSV, Institut de Génomique, Genoscope, 2 rue Gaston Crémieux, F-91057 Evry, France

[5] CNRS-UMR 8030 F-91057 Evry, France.

[6] UEVE, Université d'Evry, F-91057 Evry, France

Running title : PAH dioxygenases from uncultured soil bacteria

*Corresponding author:

Yves Jouanneau

CEA/iRTSV/CBM

F-38054 Grenoble Cedex 9, France.

Tel. : 33 438 78 43 10 ; Fax : 33 438 78 54 87

e-mail : yves.jouanneau@cea.fr

ABSTRACT

Ring hydroxylating dioxygenases (RHDs) play a crucial role in the biodegradation of a range of aromatic hydrocarbons found on polluted sites, including polycyclic aromatic hydrocarbons (PAHs). Current knowledge on RHDs comes essentially from studies on culturable bacterial strains while compelling evidence indicates that pollutant removal is mostly achieved by uncultured species. In this study, a combination of DNA-SIP labeling and metagenomic sequence analysis was implemented to investigate the metabolic potential of main PAH degraders on a polluted site. Following *in situ* labeling using ^{13}C-phenanthrene, the labeled metagenomic DNA was isolated from soil and subjected to shotgun sequencing. Most annotated sequences were predicted to belong to Betaproteobacteria, especially Rhodocyclaceae and Burkholderiales, consistent with previous findings showing that main PAH degraders on this site were affiliated to these taxa. Based on metagenomic data, four RHD gene sets were amplified and cloned from soil DNA. For each set, PCR yielded multiple amplicons with sequences differing by up to 321 nucleotides (17%), reflecting the great genetic diversity prevailing in soil. RHDs were successfully overexpressed in *E. coli*, but full activity required the co-expression of two electron carrier genes, also cloned from soil DNA. Remarkably, two RHDs exhibited much higher activity when associated with electron carriers from a Sphingomonad. The four RHDs showed markedly different preferences for 2- and 3-ring PAHs, but were poorly active on 4-ring PAHs. Three RHDs preferentially hydroxylated phenanthrene on the C-1 and C-2 positions rather than on the C-3, C-4 positions, suggesting that degradation occurred through an alternate pathway.

INTRODUCTION

Bioremediation procedures used to treat polluted sites rely on specialized microorganisms that can transform or utilize organic pollutants as carbon sources. Knowledge on pollutant biodegradation mainly arises from studies on pure strains that have been isolated from contaminated sites. For example, numerous bacterial strains able to degrade PAHs have been used to elucidate relevant degradation pathways and characterize some of the enzymes involved (1, 2). Nevertheless, exploration of the diversity of soil bacteria using culture-independent molecular techniques revealed that soils contain a great taxonomic richness and established that bacterial isolates described so far represented no more than 5% of the bacterial diversity (3). As a consequence, it could be anticipated that bacteria responsible for PAH removal *in situ* would be largely unknown, and would differ from previously studied isolates. Accordingly, Sphingomonads detected on polluted sites by 16S rRNA sequence analysis were found to be different from described species in this taxonomic group (4). In the last decade, the implementation of stable isotope probing (SIP) to track PAH degraders led to the discovery of new bacteria with interesting biodegradation potential (5, 6). Moreover, SIP approaches also revealed that most PAH-degrading bacteria identified in contaminated soils were affiliated to uncultured microorganisms (5, 7-10). Notably, Betaproteobacteria were shown to form a dominant subgroup of the phenanthrene degrading community found in polluted soils, suggesting that they played a major role in PAH degradation in soil (7, 8). Specifically, soil bacteria utilizing phenanthrene included several taxa related to Burkholderiales, as well as unclassified Rhodocyclaceae. Closely related representatives of the latter family have been found in contaminated soils from America (9), Europe (7) and a tropical region of Africa (11). A Rhodocyclaceae member appeared as the main bacterium in a consortium obtained by enrichment from soil after repeated cultivation on a pyrene-containing minimal medium (12). Although the bacterium could not be isolated in pure

culture, a metagenomic analysis of the DNA isolated from a simplified consortium consisting of the Rhodocyclaceae member and 3 other detectable bacterial species, gave insights into their metabolic capabilities. Eight sets of genes coding for ring-hydroxylating dioxygenases were identified in separate contigs. RHDs are multi-component metalloenzymes, which catalyze the first step in the bacterial degradation of various aromatic hydrocarbons. PAH-specific RHDs were found to fall into two families based on phylogenetic comparison of available sequences in public databases (13-15). Six of the RHD enzymes mentioned above were cloned and shown to catalyze the hydroxylation of several PAHs including pyrene (12).

The goal of the present study was to learn on the metabolic potential of PAH degraders in a polluted soil by combining DNA-SIP with metagenomics. Although this combination has been recognized as a promising new approach in soil bioremediation studies (16), it has not been frequently implemented so far (17). In this work, we investigated the soil bacterial community of a facility collecting the road runoffs of a highway. A SIP analysis of phenanthrene-utilizing bacteria in this soil previously demonstrated the preponderance of Betaproteobacteria, especially members of the *Acidovorax, Rhodoferax* and *Hydrogenophaga* genera, as well as unclassified Rhodocyclaceae (7). Moreover, a PCR-based analysis of the diversity of RHDs associated with phenanthrene degradation in the same soil revealed the occurrence of five groups of enzymes, three of which were poorly related to known dioxygenases, with sequence identities in the 60-80% range with best matches in databases (18). To get further information on phenanthrene degradation in soil, we have undertaken a metagenomic analysis involving a scaled-up SIP experiment in order to isolate enough labeled DNA for subsequent shotgun sequencing. From the resulting metagenomic data, four sets of RHD genes were cloned and overexpressed in *Escherichia coli*. The RHDs were found to be distantly related to the enzymes of known bacterial isolates but shared high similarities

with the RHDs found in the pyrene-degrading consortium mentioned above. The catalytic properties of the enzymes with respect to the oxidation of 2- to 4-ring PAHs were determined.

MATERIALS AND METHODS

Bacterial strains and plasmids

Bacterial strains and plasmids used in this study are listed in Table 1. *E. coli* strains were grown in rich medium (Luria-Bertani) at 37°C with appropriate antibiotics as previously described (19).

SIP experiments and isolation of labeled DNA from soil

Soil sampling was carried out on October 29, 2010, on a previously described study site, which is designed to collect road runoffs from a highway (7). SIP experiments were conducted in 250-ml microcosms containing 20 g of soil. Briefly, the soil of each microcosm was mixed with 2.5 mg of ^{13}C-phenanthrene dissolved in 0.25 ml DMSO. Microcosms were closed with rubber stoppers, then incubated in the dark for 6 days at 25°C under static conditions. Microcosms that did not receive phenanthrene were incubated under identical conditions to serve as controls. Mineralization of the labeled substrate was monitored by GC/MS quantification of $^{13}CO_2$ in the gas phase (7). DNA was extracted from 10-g lots of wet soil using the PowerMax soil DNA extraction kit (Mo Bio laboratories), then separated by CsCl isopycnic ultracentrifugation as previously described (7). The heavy fractions of the gradient containing ^{13}C DNA were subjected to PCR tests as illustrated in figure S1. The primer pair employed was previously shown to amplify a 950-bp fragment of the RHD alpha subunit genes from soil Betaproteobacteria (RHD-Beta-Grp1f and RHD-Beta-Grp1r) (18). Fractions that responded positively to the PCR test were pooled from repeated preparations

120 involving 18 soil DNA extractions followed by gradient fractionation, yielding a total of 30,6
121 µg of labeled DNA.
122
123 **Metagenomic DNA sequencing and analysis**
124 Sequencing was carried out using 454 pyrosequencing (Roche Biosciences) as well as
125 Illumina technology. A 8-Kb paired-end library was constructed according to the protocol for
126 the 454 Titanium apparatus. Two sequencing runs were performed generating 784.5 Mbp of
127 raw data. A library of 362-bp inserts (average size) was constructed according to the Illumina
128 HiSeq 2000 protocol. Six lanes of 100 bp pair-end sequencing were used to generate 130.4
129 Gbp of raw data.
130 The Titanium sequences were assembled by Newbler (version MapAsmResearch-04/19/2010-
131 patch-08/17/2010) and the sequences of the scaffolds were corrected using the Illumina
132 sequences (20). Resulting assembly was composed of 69,435 contigs (11,909 of them being
133 larger than 500 bp, summing up to 9,014,532 bp) organized in 288 scaffolds for a cumulative
134 scaffold size of 824,483 bp. Contigs larger than 500 bp were scanned for 16 S ribosomal
135 genes by BLASTn similarity search against the Greengenes otu97 database (release 13-5)
136 (21). Contigs with significant matches (> 90 % nucleic identity over at least 300 bp) were
137 retained for subsequent analysis. Selected sequences were also compared to entrees in the
138 NCBI nucleotide database using BLASTn.
139 Predictions of coding regions from the obtained set of contigs were performed with
140 MetaGeneAnnotator with default parameters (22), resulting in a total of 85,156 coding
141 sequences. Predicted protein sequences were compared with the UniProtKB database (release
142 of 2014/01/24) using the LASSAP implementation of the BLASTp algorithm with a threshold
143 e-value of $1e^{-5}$ (23).
144

Cloning of RHD-encoding genes

PCR amplifications of selected RHD-encoding genes were carried out on a Tpersonal thermocycler (Whatman Biometra), using primer pairs depicted in Table 2 with unfractionated DNA from phenanthrene-spiked soil as template. Reactions were performed in a 25- or 50-µl total volume containing 1x polymerase buffer, 1.5 mM $MgSO_4$, 0.2 mM of each dNTP, 0.3 µM of each primer, 2 ng/µl of metagenomic DNA, 0.02 U/µl of high fidelity DNA polymerase, 40 ng/µl of phage T4 gp32 (New England Biolabs). The KOD Hot Start DNA polymerase was most commonly used (Merck Novagen), under the following PCR conditions: DNA denaturation at 95°C for 2 min, then 30 cycles of denaturation at 95°C for 20s, annealing for 15s at the temperature adequate to the chosen primer pair, and extension at 72°C for 30s/kb. For the amplification of the *pahAa* and *pahAb* genes, a touch down program was implemented where the annealing temperature was lowered from 53 to 49°C by 1° increments during the first 5 cycles of the PCR. Occasionally, the KOD enzyme was replaced by the Q5 high fidelity DNA polymerase (New England Biolabs) with the following modifications: the extension step was increased to 1 min/kb, and a final 2-min extension was added at the end of the PCR program. PCRs where template DNA was from soil samples not supplemented with phenanthrene were run as controls.

PCR products were purified by agarose gel electrophoresis followed by DNA fragment extraction with the NucloSpin Extract II kit (Macherey-Nagel), then cloned using the CloneJET PCR cloning kit (Thermo Scientific). Plasmid inserts were sequenced on both strands by Eurofins MWG/Operon (Germany). DNA sequences obtained were analyzed using the ApE software available at http://biologylabs.utah.edu/jorgensen/wayned/ape/ and compared to those in databases using BLASTn available on the EMBL EBI website. Search for proteins similar to our translated sequences was performed in the UniProt knowledgebase

169 using BLASTp. Neighbor-joining analysis of RHD alpha subunit sequences was done on the
170 phylogeny.fr website using the Oneclick option (24).
171
172 **Construction of plasmids for the overexpression of RHD genes in *E. coli***
173 RHD genes were cloned into plasmid pET15b and overexpressed in strain BL21(DE3). The
174 gene pairs *pahAc2Ad2* and *pahAc8Ad8* were subcloned as NdeI-XhoI fragments from
175 plasmids pJCA10 and pJCA14 into pET15b to give plasmids pCAE1 and pCAE4,
176 respectively. Since the *pahA4Ad4* sequence contains an internal NdeI site, a 2-kb fragment
177 carrying the 2 genes was recovered after partial digestion of pJCA9 with NdeI and XhoI, and
178 subsequently cloned into pET15b to give pCAE5. For subcloning of *pahAc5Ad5*, a 1.8 kb
179 fragment carrying these 2 genes was amplified with primers C451F2 and JCA7-R2 (Table 2)
180 using pJCA7 as template, then the PCR product was digested by NdeI and XhoI before being
181 cloned into pET15b to give pCAE7.
182 Plasmid pCAZ1 carrying the *pahAa* and *pahAb* genes, which encode a NAD(P)H
183 oxydoreductase and a ferredoxin, respectively, was constructed in two steps. First, the *pahAb*
184 sequence was isolated from pJCA2 as a XhoI-NcoI fragment, then cloned downstream of
185 *pahAa* in pJCA3 digested by SalI and NcoI. The resulting plasmid, called pJCA5, was cut by
186 XhoI and XbaI, to isolate a 1.5-kb fragment encompassing the 2 genes, which was cloned into
187 pIZ1036 digested by SalI and XbaI, to give pCAZ1, where the two genes are under the
188 control of the tac promoter (Table 1).
189
190 **RHD overproduction and assays**
191 Gene overexpression was performed in strain BL21(DE3), which had been co-transformed
192 with pCAZ1 and one of the pET15b derivatives carrying a pair of RHD genes. In some
193 experiments, plasmid pCAZ1 was replaced by pIBA34, thus allowing the expression of an

194 alternate pair of electron carriers. Recombinant *E. coli* strains were grown at 37°C in LB
195 medium until the optical density reached around 1.0 at 600 nm (OD_{600}). Then, cultures were
196 induced with 0.5 mM IPTG and further incubated for 21-22 h at 25°C under orbital shaking at
197 180 rpm. Cells were harvested by centrifugation, then washed and resuspended to an OD_{600} of
198 2.0 in the minimal M9 medium containing 10 mM glucose. For PAH oxidation assays, the
199 suspension was distributed into 2-ml Eppendorf tubes (1 ml/tube), which had received 1.0
200 µmole of chosen PAH, applied as a 20 mM stock solution in acetone. Three replicates per
201 PAH were prepared. Cells were incubated for 6 h (naphthalene, biphenyl, phenanthrene) or 24
202 h (anthracene, fluorene, pyrene, fluoranthene) at 25°C with vigorous shaking.. Then, cell
203 suspensions were centrifuged, and the supernatant fluid was extracted with an equal volume
204 of ethyl acetate in the presence of 10 µM 2,3-dihydroxybiphenyl (Sigma-Aldrich), added as
205 an internal standard. Dried extracts were taken up in 0.2 ml acetonitrile and analyzed by
206 GC/MS as *n*-butylboronate (NBB) derivatives as previously described (25). Dihydrodiols
207 were quantified using calibration curves obtained by analyzing samples of purified
208 naphthalene 1,2-dihydrodiol or phenanthrene 3,4-dihydrodiol in the 5-100 µM range (26), and
209 normalized with respect to the internal standard concentration. Pyrene oxidation products
210 were instead acetylated by treating dried extracts with 40 µl pyridine and 60 µl acetic
211 anhydride for 30 min at 60°C. Samples were analyzed by GC-MS in single ion monitoring
212 mode (with 260 and 320 as selected m/z), using purified pyrene 4,5-dihydrodiol as a standard
213 for calibration. The pyrene dihydrodiol as well as other diols mentioned above were prepared
214 as previously described (27). Activities are expressed as the amount of dihydrodiol formed
215 (µM) per hour per ml of culture, normalized to a density of 1.0 (OD_{600}). Fluorene oxidation
216 products were also analyzed after trimethylsilylation as previously described (26).
217
218 **SDS-PAGE analysis**

Expression of recombinant RHDs in *E. coli* was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on 12.5% slab gels using a SE 260 Mighty Small II system (Hoefer™). Gels were processed and stained as previously described (26).

Nucleotide sequence accession numbers: the gene sequences described in this study have been deposited in the European Nucleotide Archive database under accession numbers HG918050-HG918067

RESULTS

Overview of sequence predictions from the genomic DNA of soil phenanthrene degraders

The first goal of this study was to isolate enough genomic DNA from soil bacteria able to degrade PAHs in order to afford subsequent shotgun sequencing. A large-scale labeling experiment with ^{13}C-phenanthrene as probe was implemented, consisting of several SIP experiments performed according to a procedure that enabled us to identify phenanthrene degraders in the soil of the same study site (7). In soil microcosms, labeled phenanthrene underwent rapid metabolization after a 2-d lag period, as shown by recording the time course of mineralization (Fig. S2). At day 6, DNA was extracted from soil, then subjected to isopycnic separation by repeated CsCl gradient centrifugations. For each run, heavy fractions containing the labeled DNA were selected on the basis of a PCR test as depicted in Fig. S1. The primers used in this test targeted phenanthrene-specific RHD genes previously identified by SIP in the same soil (18), The fractions that responded positively to this PCR test most likely contained DNA from soil PAH-degrading bacteria. Pooling the fractions resulted in a total of 30.6 µg of DNA, which was subjected to both Roche 454 and Illumina sequencing.

Analysis of metagenomic data allowed the assembly of 69,435 contigs, including 11,909 contigs longer than 500 bp, the longest comprising 8050 bp. Sequence annotation revealed that most open reading frames (ORFs) were partial as a consequence of the short length of the contigs. Nevertheless, a total of 85,156 coding sequences were deduced from the metagenomic data, including 63,802 sequences that showed at least one match in the UniprotKB database. About 48% of these protein sequences were predicted to belong to Betaproteobacteria (Table S1). In addition, 11 sequences coding for 16S rRNA subunits were identified, four of which were affiliated to Betaproteobacteria (Tablr 3). These results are consistent with previous findings showing that dominant PAH degraders in the same soil

252 belonged to this bacterial class (7). Moreover, seven 16S rRNA gene sequences matched
253 OTUs previously affiliated to members of the Gamma- and Betaproteobacteria, which likely
254 contributed greatly to phenanthrene degradation, including two sequences from uncultured
255 Rhodocyclaceae (Tablr 3),
256 Approximately 510 sequences were annotated as components of dioxygenases possibly
257 involved in the biodegradation of aromatic hydrocarbons (Fig. S3). One half of them would
258 correspond to alpha or beta subunits of RHDs with undefined substrate specificity, whereas
259 the other half included benzoate dioxygenases and enzymes involved in lower steps of the
260 hydrocarbon metabolic pathways, such as phthalate or protocatechuate dioxygenases, and
261 extradiol dioxygenases. Only a few sequences coding for RHD-associated electron carriers
262 were identified, suggesting that the oxygenase components of many of the three-component
263 RHDs might share common electron carriers. In this study, we focused on the RHD
264 sequences found among the 600 longest contigs derived from our metagenomic analysis
265 (Table S2)
266
267 **Cloning and sequence analysis of four RHD genes and one set of genes coding for related**
268 **electron carriers**
269 Contigs 204, 332, 341, 427, 451 and 569 contained ORFs coding potentially for RHD alpha
270 subunits, four of which also showed an ORF for a beta subunit (Table S2). The α and β RHD
271 subunits deduced from the contig 204 sequence showed highest similarities (≈ 60% identity)
272 with the subunits of a RHD annotated as a phenylpropionate dioxygenase in
273 *Pseudoxanthomonas spadix,* a BTEX-degrading Gammaproteobacterium (28). The α subunit
274 also showed a similar relatedness (60% identity) with the subunit of an enzyme characterized
275 as a salicylate hydroxylase in the PAH-degrading strain *Sphingomonas* CHY-1 (29). In the
276 other five contigs, the identified ORFs were similar to recently described RHD genes, which

277 were obtained from a pyrene-degrading consortium dominated by uncultured Rhodocyclaceae
278 (12). Since the contigs contained partial RHD sequences or lacked one of the subunit genes,
279 we searched in the whole set of contigs for missing sequences using relevant homologous
280 genes described in the above study as queries. Sequences corresponding to the missing 5'-end
281 of the α subunit genes present in contigs 332, 341 and 427 were found in contigs 763, 55951,
282 and 5241, respectively. For the alpha subunit genes present in contigs 451 and 569, candidate
283 beta subunits were found in contigs 3411 and 2271, respectively. Using appropriate DNA
284 primers designed after the extremities of each reconstituted pair of RHD genes (Table 2) and
285 metagenomic DNA from phenanthrene-spiked soil as template, PCR products of
286 approximately 2.0 kb were obtained in all cases except for the reaction with primers C569-F
287 and C2271-R . Sequence analysis of the cloned PCR products revealed that the RHD genes
288 were very similar but not identical to those determined by metagenomic sequencing (Table 4).
289 Moreover, analysis of another amplicon from each PCR yielded a sequence that was different
290 from that of the first clone, and from that in the relevant contig. The number of mismatches
291 between amplicons cloned from the same PCR varied from 12/1955 (0.6% for pJCA13 versus
292 pJCA14) to 321/1906 (17% for pJCA6 versus pJCA7). These results most likely reflect the
293 great diversity of gene sequences within each type of RHD in soil bacteria.
294 Contig 095 contained one gene coding for a NAD(P)H-ferredoxin oxidoreductase as well as
295 three ORFs very similar in sequence and arrangement to a gene cluster previously described
296 for a pyrene-degrading consortium in the study mentioned above (12). The reductase gene,
297 designated as *pahAa* in that study, is part of a gene cluster including *pahAb* located about 3 kb
298 upstream and encoding a ferredoxin (Fig. 1). A gene closely similar to *pahAb* was found in
299 contig 113, which also contained two genes encoding putative transcriptional regulators of the
300 LysR type (best match: Q12EV9 ; 49% identity) and MarR type (best match: D6CQZ4 ; 51%
301 identity), respectively (Fig. 1). These genes showed an opposite orientation with respect to

302 *pahAb,* and were followed by a gene similar to that coding for a cytochrome B561 in
303 S*ideroxydans lithotrophicus* (D5CTW7 ; 54% identity).
304 The two genes homologous to *pahAa* and *pahAb* were separately amplified with DNA from
305 phenanthrene-spiked soil as template. Cloned sequences showed 29 and 5 mismatches when
306 compared to those found in contigs 95 and 113, respectively. For the sake of clarity, these
307 genes, as well as the RHD genes described above, were given the same names as their
308 counterparts previously found in a pyrene-degrading consortium (Table 4). It is to note that
309 control PCRs where template DNA was from soil incubated without phenanthrene, yielded no
310 detectable product, except for the *pahAc8/pahAd8* gene pair (data not shown). This result
311 provides further evidence that the RHD genes were from main degraders that developed at the
312 expense of added phenanthrene during the six-day incubation in microcosms.
313
314 **Heterologous expression of four RHDs and electron carrier preference**
315 For each set of RHD genes identified in this work (Table 4), we selected one of the cloned
316 sequences for subcloning in plasmid pET15b and subsequent expression in *E. coli*
317 BL21(DE3) (see Table 1 and Materials and Methods for details). Hence, the *pahAc2/Ad2*,
318 *pahAc4/Ad4, pahAc5/Ad5* and *pahAc8/Ad8* genes cloned in pCAE1, pCAE5, pCAE7 and
319 pCAE4 were from pJCA10, pJCA9, pJCA7 and pJCA14, respectively. Another plasmid
320 called pCAZ1 (Table 1) was constructed and introduced in BL21(DE3) to co-express the
321 *pahAa* and *pahAb* genes coding for two electron carriers, which were assumed to associate
322 with RHD oxygenase components to form active enzyme complexes. Under appropriate
323 induction conditions, recombinant *E. coli* strains overproduced two polypeptides of around 50
324 and 20 kDa, as shown by SDS-PAGE analysis of whole cell extracts (Fig. 2). The observed
325 polypeptide sizes were consistent with those deduced from gene sequences. In this respect,
326 the PahAd5 product (184 amino acids, expected mass: 21,331 Da) showed a slower migration

327 profile than those of PahAd2 (19,870 Da), PahAd4 (19,821 Da) and PahAd8 (21,186 Da).
328 However, the observed mobility shift of PahAd5 cannot be only explained by its higher
329 molecular mass. In fact, it might be due to its basic character (theoretical isoelectric point
330 around 9.0), whereas the other beta subunits were predicted to be acidic (pI: 5.63 and 6.40 for
331 PahAd4 and PahAd8) or near neutral (pI: 7.8 for PahAd2). The overexpression of *pahAa* and
332 *pahAb* yielded polypeptides with apparent M_r of ≈ 37,000 and 13,000, in fairly good
333 agreement with the expected masses of 37,385 and 11,635 Da, as deduced from their
334 respective sequences (Fig. 2, lane 6).
335 All four oxygenases converted naphthalene to *cis*-1,2-dihydroxy 1,2-dihydronaphthalene on
336 condition that the PahAa and PahAb proteins were simultaneously produced in *E. coli*. No or
337 negligible activity was detected in strains overexpressing the oxygenase components alone (
338 Table 5). The activity of PahAc4/Ad4 was about one order of magnitude lower compared to
339 the other oxygenases, although its expression level in *E. coli* was similar as judged from SDS-
340 PAGE (Fig. 2). This suggested that either naphthalene was not a good substrate for this
341 enzyme or the electron carriers were inadequate. The second hypothesis was tested by
342 replacing pCAZ1 by a plasmid (pIBA34) overexpressing PhnA4 and PhnA3, the electron
343 carriers associated to the naphthalene dioxygenase from *Sphingomonas* CHY-1 (19, 26). The
344 resulting strain BL21(DE3)(pCAE5)(pIBA34) exhibited a dioxygenase activity 20-fold higher
345 than that measured in the strain expressing PahAa and PahAb from pCAZ1 (Table 5).
346 Assuming that the two compared strains showed equivalent expression levels of PahAc4/Ad4
347 (data not shown), and similar levels of the electron carriers since pCAZ1 and pIBA34 were
348 derived from the same plasmid, it is concluded that the higher activity found in the strain
349 harboring pIBA34 most likely reflected a better compatibility of the oxygenase with the
350 PhnA4/PhnA3 electron carriers. Likewise, the PahAc2/Ad2 oxygenase showed a higher
351 activity with PhnA4/PhnA3 than with PahAa/PahAb, although the activity ratio was only 1.5

352 in that case. These results suggested that the latter two oxygenases might operate in cells
353 related to Sphingomonads, consistent with their sequence similarity with PAH dioxygenases
354 from this taxonomic group (see Fig. 3 and discussion below). From the the metagenomic data
355 we identified one ferredoxin-encoding gene in contig 9967, whose closest match was a *bphA3*
356 gene from *Novosphingobium* sp. PCY (Accession No. S5YUI9). However, at the protein
357 level, sequence similarities with BphA3 from strain PPIY (53% identity) or with PhnA3 from
358 strain CHY-1 (51% identity) were relatively low, thus precluding a possible affiliation of that
359 protein to any taxon.
360
361 **Substrate specificity towards 2- to 4-ring PAHs**
362 The four recombinant RHDs exhibited a narrow specificity for 2- and 3-ring PAHs,
363 naphthalene being the preferred substrate (Table 6). PahAc8/Ad8 showed the broadest
364 substrate range and oxidized biphenyl at about the same rate as naphthalene. The enzyme also
365 oxidized 3-ring PAHs at a significant rate, given that the relatively small amount of
366 dihydrodiol detected with fluorene as substrate reflected only part of the activity. Indeed,
367 analysis of trimethylsilylated fluorene oxidation products generated by PahAc8/Ad8, allowed
368 the detection of one monohydroxy- (M+ = 254), and three dihydroxy-fluorene (M+ = 342) by
369 GC/MS (data not shown). PahAc5/Ad5 appeared as the only RHD able to generate
370 measurable amounts of dihydrodiol from pyrene (Table 6), although the dihydroxylation rate
371 was low compared to that observed with naphthalene (≈ 0.054%). With phenanthrene as
372 substrate, the enzyme produced two dihydrodiols with a 1:2 molar ratio, the isomer
373 hydroxylated on the C-3 and C-4 positions being the less abundant. The predominant isomer,
374 showed a retention time (14.63 min versus 14.27 min for the 3,4 isomer) and a mass spectrum
375 distinct from those of the previously identified 9,10-isomer (30). Based on comparisons with
376 previously published data (31), we assumed that the second dihydrodiol formed by

PahAc5/Ad5 was hydroxylated on C-1 and C-2 positions. Likewise, the PahAc2/Ad2 and PahAc4/Ad4 oxygenases formed two isomers from phenanthrene with an even higher proportion of the 1,2-isomer (Table 6). Despite small differences in their specific activity, the two enzymes, which share high sequence similarity (81-84% identity overall), exhibited the same narrow selectivity for naphthalene and phenanthrene.

DISCUSSION

Combining SIP with shotgun sequencing appears as a valuable strategy for targeting uncultured microorganisms with desired metabolic functions and for extracting relevant genetic information from soil DNA. As shown in the present study, this approach was successfully employed to specifically investigate the potential of PAH-degrading bacteria that predominate in contaminated soil, resulting in a large body of genomic sequences from which interesting new biocatalysts have been cloned and functionally characterized. Ideally, metagenomic data should give upon assembly large contigs representing portions of bacterial chromosomes bearing clusters of genes involved in the same metabolic function, such as pollutant degradation. In the present work, the limited size of the contigs precluded the deciphering of large gene clusters related to PAH degradation. The limited length of the contigs might primarily be a consequence of the complexity of the recovered metagenomic sequences, likely originating from multiple genomes. Despite the sieving effect of SIP, which removed the DNA from bacteria unrelated to phenanthrene metabolism from the analysis, the detection of ten 16S rRNA gene sequences in our metagenomic data indicated that the labeled DNA contained fragments of at least as many genomes. In addition, there seems to be a high degree of polymorphism among some isofunctional genes present in soil bacteria, as exemplified by the fact that cloned amplicons specific for each RHD gene set were different in sequence. This polymorphism might have hindered the assembly of short reads generated

by 454 pyrosequencing or led to artifacts due to mosaic assembly. Another possible cause of the contig shortness might be the depth of the sequencing effort, which was limited by the scarcity of labeled DNA recovered from soil. To overcome the problem associated wih the low recovery of labeled DNA inherent to SIP, amplification methods such as multiple displacement amplification may be used. Employing this method, Wang et al. were able to assemble a *nag* gene cluster responsible for the conversion of naphthalene to salicylate (17). Presumably, this gene cluster was part of the genome of an uncultured *Acidovorax* sp, which was identified as a prevalent naphthalene degrader *in situ*.

Although partial and fragmented, the information derived from our metagenomic data provides new and valuable insights into the metabolic potential of soil phenanthrene degraders. Based on sequence annotation, a majority of genes would belong to Betaproteobacteria, one fourth of which appear as Rhodocyclaceae. Even though some of the annotations might be wrong, the high proportion of genome sequences related to Rhodocyclaceae is corroborated by the occurrence of two (of 11) 16S rRNA sequences affiliated to this family in the metagenomic DNA (Table 3). The two 16S rRNA sequences of interest are identical to prevalent ones previously detected by DNA-SIP in the PAH degrading community of the same soil (OTU17 and 101), (7). In addition to Rhodocyclaceae, our metagenomic data also underscored the importance of Gammaproteobacteria, as indicated by the occurrence of four 16S rRNA sequences, which were representative of two unclassified bacterial groups previously referred to as OTU2 and OTU153 (Table 3, (7)). Species represented by OTU2 are phylogenetically related to soil bacteria designated as Pyr group 2, first identified as pyrene degraders (9), and later shown to also degrade other 4-ring PAHs (32). On the other hand, bacteria of the *Acidovorax, Rhodoferax* and *Hydrogenophaga* genera, which were previously identified as prevalent phenanthrene degraders in the studied soil, were poorly represented in the annotations of the sequence data. This observation might

reflect some variability of the soil bacterial community, which possibly underwent changes during the one-year period separating soil sampling for this work and our previous study (7).

As a first step towards a better understanding of the metabolism of soil phenanthrene degraders, we have cloned and characterized four RHDs responsible for the initial attack of PAHs. The enzymes exhibited limited sequence similarity (50-70% identity) with well characterized PAH dioxygenases described so far, which were isolated from culturable strains (13). A phylogenetic comparison of selected PAH dioxygenases indicated that three enzyme sequences (PahAc2, PahAc4 and PahAc5) clustered with poorly characterized RHDs from Sphingomonads (AhdA1a or BphA1a; Fig. 3). On the other hand, the PahAc8 sequence appeared as distant from NahAc/NagAc from Beta- and Gammaproteobacteria as from PhnAc from Betaproteobacteria. PahAc3, which could not be biochemically studied in this work, was more closely related to the PhnAc-type of enzymes.

The PahAc sequences also differed from RHDs previously identified by a PCR method used to amplify a region coding for the catalytic domain (314 residues) of the alpha subunit and isolated from the same soil (18). In that study, retrieved RHD sequences were grouped in five main clusters, two of which were related to PAH dioxygenases found in Betaproteobacteria, and one other cluster to Alphaproteobacteria (mainly Sphingomonads). One sequence was almost identical to the corresponding part of PahAc8. The other three RHDs have sequences that do not match any of the previously detected sequences, perhaps because their bacterial hosts were absent due to changes in the soil bacterial community. Alternatively, the primer pairs we used in our previous study to amplify RHD alpha subunits might have been inadequate to detect *pahAc2, pahAc4*, and *pahAc5*. In this respect, sequence analysis predicts that the reverse primers employed in our previous work would not correctly hybridize with the coding sequences of the latter three genes. On the other hand, *pahAc3* present in contig 569 (Table S2) is closely similar to numerous sequences affiliated to Betaproteobacteria and

452 referred to as cluster 4 (18). Unfortunately, we have been unable to clone a pahAc3 gene with
453 its associated beta subunit gene, precluding the functional characterization of the
454 corresponding RHD.
455 Finally, the RHDs that best matched those described in this work were recently obtained from
456 a bacterial consortium selected by enrichment with pyrene as carbon source (12). Eight RHDs
457 have been cloned from this pyrene-specific consortium, six of which were apparently able to
458 attack 2- to 4-ring PAHs to various extent. It has been proposed that all these enzymes
459 belonged to the same bacterial type, related to Rhodocyclaceae, as it appeared as a
460 predominant member of the consortium. Curiously, all eight RHDs except one, appear to have
461 counterparts in the phenanthrene degrading community examined in this work, since, in
462 addition to the four studied RHDs, sequences similar to *pahAc1, pahAc3* and *pahAc6* were
463 detected in contigs 38041, 569 and 1214, respectively. Our data also suggest that they might
464 have the same bacterial source. Examples of bacterial isolates with multiple functional PAH
465 dioxygenases are rare, especially among Betaproteobacteria (13). In Sphingomonads, up to
466 six sets of RHD-like sequences have been described (33, 34), but only one set appeared to
467 encode a PAH dioxygenase (19), the other enzymes catalyzing the hydroxylation of salicylate
468 and methyl salicylates (29, 35). In *Mycobacteria*, some strains have been shown to synthesize
469 up to three types of PAH dioxygenases with distinct specificities (30, 36).
470 The substrate range of the four RHDs is relatively narrow and limited to 2 and 3-ring PAHs.
471 Only one RHD proved capable of utilizing pyrene, which contrasts with a previous report
472 providing evidence that similar enzymes could degrade the 4-ring substrate (12). Also, while
473 PahAc8/Ad8 exhibited the broadest substrate specificity, its counterpart called RHD-8
474 showed insignificant activity with any PAH except phenanthrene and pyrene. Discrepancies
475 might be due primarily to differences in experimental conditions, although differences in
476 amino acid sequence between homologous RHDs should also be taken into consideration.

Percent identity between homologous RHD components in the present and the cited study vary between 77.5 and 94% (Table 4). Marked changes in specificity were observed between PAH dioxygenases from Sphingomonads showing equivalent sequence relatedness. Notably, the RHD from *Spingobium yanoikuyae* B1 has a preference for biphenyl (37), whereas the RHDs from *Sphingomonas* CHY-1 and LH128 utilize naphthalene as best substrate (26, 38). Also, biphenyl dioxygenases displaying as high as 99% sequence identity proved markedly different in substrate specificity toward polychlorobiphenyls (39). Hence, compared to the enzymes described in this work, the RHDs described by Singleton et al. might have a better activity for pyrene because of the bacterial enrichment on this PAH, which preceded RHD gene isolation (12).

When incubated with phenanthrene, three of the studied RHDs generated more 1,2-dihydrodiol than 3,4-dihydrodiol (Table 6), a rather unexpected result since the latter compound is thought to be the most common intermediate in bacterial degradation pathways (2). Although dioxygenation of phenanthrene on the C-1 and C-2 positions has been shown to occur in *Sphingomonas* P2 (40) and in *Burkholderia* sp. C3 (41), and gives rise to effective metabolization, most known RHDs do not generate significant amounts of 1,2-dihydrodiol from phenanthrene. In this respect, a mutant form of the naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB9816-4 appears as an exception. Its ability to form an excess of phenanthrene 1,2-dihydrodiol resulted from the replacement of the Phe 352 residue by a valine at the enzyme active site (31). The RHDs described in the present study have a phenylalanine in equivalent position, indicating that their ability to better hydroxylate phenanthrene on the C-1 and C-2 positions is not due to a similar amino acid substitution. Our work provides evidence that this catalytic property might be a common feature of PAH dioxygenases from soil bacteria, suggesting that it could confer a selective advantage to phenanthrene degraders.

Acknowledgments

The authors are grateful to the CNRS and Univ Grenoble Alpes for funding. This work was also supported in part by a grants from the Rhône-Alpes region and from the ANR "programme Labex" (ARCANE project n° ANR-11-LABX-003).

References

1. **Doyle E, Muckian L, Hickey AM, Clipson N.** 2008. Microbial PAH degradation. Adv. Appl Microbiol **65:**27-66.

2. **Peng RH, Xiong AS, Xue Y, Fu XY, Gao F, Zhao W, Tian YS, Yao QH.** 2008. Microbial biodegradation of polyaromatic hydrocarbons. FEMS Microbiol Rev **32:**927-955.

3. **Cole JR, Konstantinidis K, Farris RJ, Tiedje JM.** 2010. Microbial diversity and phylogeny: extending from rRNAs to genomes, p. 1-19. *In* Liu W-T, Jansson JK (ed.), Environmental molecular microbiology. Caister Academic Press, Norfolk, UK.

4. **Leys NMEJ, Ryngaert A, Bastiaens L, Verstraete W, Top EM, Springael D.** 2004. Occurrence and phylogenetic diversity of *Sphingomonas* strains in soils contaminated with polycyclic aromatic hydrocarbons. Appl. Environ. Microbiol. **70:**1944-1955.

5. **Jeon CO, Park W, Padmanabhan P, DeRito C, Snape JR, Madsen EL.** 2003. Discovery of a bacterium, with distinctive dioxygenase, that is responsible for in situ biodegradation in contaminated sediment. Proc. Natl. Acad. Sci. U. S. A. **100:**13591-13596.

6. **Singleton DR, Ramirez LG, Aitken MD.** 2009. Characterization of a polycyclic aromatic hydrocarbon degradation gene cluster in a phenanthrene-degrading *Acidovorax* Strain. Appl. Environ. Microbiol. **75:**2613-2620.

7. **Martin F, Torelli S, Le Paslier D, Barbance A, Martin-Laurent F, Bru D, Geremia R, Blake G, Jouanneau Y.** 2012. Betaproteobacteria dominance and diversity shifts in the bacterial community of a PAH-contaminated soil exposed to phenanthrene. Environ Pollut **162:**345-353.

8. **Singleton DR, Powell SN, Sangaiah R, Gold A, Ball LM, Aitken MD.** 2005. Stable-isotope probing of bacteria capable of degrading salicylate, naphthalene, or

phenanthrene in a Bioreactor treating contaminated soil. Appl Environ Microbiol **71:**1202-1209.

9. **Singleton DR, Sangaiah R, Gold A, Ball LM, Aitken MD.** 2006. Identification and quantification of uncultivated Proteobacteria associated with pyrene degradation in a bioreactor treating PAH-contaminated soil. Environ Microbiol **8:**1736-1745.

10. **Uhlik O, Wald J, Strejcek M, Musilova L, Ridl J, Hroudova M, Vlcek C, Cardenas E, Mackova M, Macek T.** 2012. Identification of bacteria utilizing biphenyl, benzoate, and naphthalene in long-term contaminated soil. Plos One **7**.

11. **Regonne RK, Martin F, Mbawala A, Ngassoum MB, Jouanneau Y.** 2013. Identification of soil bacteria able to degrade phenanthrene bound to a hydrophobic sorbent in situ. Environ Pollut **180:**145-151.

12. **Singleton DR, Hu J, Aitken MD.** 2012. Heterologous expression of polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenase genes from a novel pyrene-degrading Betaproteobacterium. Appl Environ l Microbiol **78:**3552-3559.

13. **Jouanneau Y, Martin F, Krivobok S, Willison JC.** 2011. Ring-hydroxylating dioxygenases involved in PAH biodegradation : structure, function and biodiversity., p. 149-175. *In* Koukkou AI (ed.), Microbial bioremediation of non metals: current research. Caister Academic Press, Norflok, UK.

14. **Kweon O, Kim S-J, Baek S, Chae J-C, Adjei M, Baek D-H, Kim Y-C, Cerniglia C.** 2008. A new classification system for bacterial Rieske non-heme iron aromatic ring-hydroxylating oxygenases. BMC Biochem. **9:**11.

15. **Nam JW, Nojiri H, Yoshida T, Habe H, Yamane H, Omori T.** 2001. New classification system for oxygenase components involved in ring- hydroxylating oxygenations. Biosci. Biotechnol. Biochem. **65:**254-263.

16. **Uhlik O, Leewis MC, Strejcek M, Musilova L, Mackova M, Leigh MB, Macek T.** 2013. Stable isotope probing in the metagenomics era: A bridge towards improved bioremediation. Biotechnology Advances **31:**154-165.

17. **Wang Y, Chen Y, Zhou Q, Huang S, Ning K, Xu J, Kalin RM, Rolfe S, Huang WE.** 2012. A culture-independent approach to unravel uncultured bacteria and functional genes in a complex microbial community. Plos One **7**.

18. **Martin F, Malagnoux L, Violet F, Jakoncic J, Jouanneau Y.** 2013. Diversity and catalytic potential of PAH-specific ring-hydroxylating dioxygenases from a hydrocarbon-contaminated soil. Appl Microbiol Biotechnol **97:**5125-5135.

19. **Demaneche S, Meyer C, Micoud J, Louwagie M, Willison JC, Jouanneau Y.** 2004. Identification and functional analysis of two aromatic ring-hydroxylating dioxygenases from a *Sphingomonas* strain degrading various polycyclic aromatic hydrocarbons. Appl Environ Microbiol **70:**6714-6725.

20. **Aury JM, Cruaud C, Barbe V, Rogier O, Mangenot S, Samson G, Poulain J, Anthouard V, Scarpelli C, Artiguenave F, Wincker P.** 2008. High quality draft sequences for prokaryotic genomes using a mix of new sequencing technologies. BMC genomics **9:**603.

21. **DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol **72:**5069-5072.

22. **Noguchi H, Taniguchi T, Itoh T.** 2008. Metageneannotator: Detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA res. **15:**387-396

581 23. **Glemet E, Codani JJ.** 1997. LASSAP, a LArge Scale Sequence compArison
582 Package. Comput Appl Biosci **13:**137-143.
583 24. **Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF,**
584 **Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O.** 2008. Phylogeny.fr:
585 robust phylogenetic analysis for the non-specialist. Nucl acids res **36:**W465-469.
586 25. **Schuler L, Chadhain SMN, Jouanneau Y, Meyer C, Zylstra GJ, Hols P, Agathos**
587 **SN.** 2008. Characterization of a novel angular dioxygenase from fluorene-degrading
588 *Spingomonas* sp strain LB126. Appl Environ Microbiol **74:**1050-1057.
589 26. **Jouanneau Y, Meyer C, Jakoncic J, Stojanoff V, Gaillard J.** 2006.
590 Characterization of a naphthalene dioxygenase endowed with an exceptionally broad
591 substrate specificity toward polycyclic aromatic hydrocarbons. Biochemistry
592 **45:**12380-12391.
593 27. **Jouanneau Y, Meyer C.** 2006. Purification and characterization of an arene cis-
594 dihydrodiol dehydrogenase endowed with broad substrate specificity toward
595 polycyclic aromatic hydrocarbon dihydrodiols. Appl Environ Microbiol **72:**4726-
596 4734.
597 28. **Lee SH, Jin HM, Lee HJ, Kim JM, Jeon CO.** 2012. Complete genome sequence of
598 the BTEX-degrading bacterium *Pseudoxanthomonas spadix* BD-a59. J Bacteriol
599 **194:**544.
600 29. **Jouanneau Y, Micoud J, Meyer C.** 2007. Purification and characterization of a
601 three-component salicylate 1-hydroxylase from *Sphingomonas* sp. strain CHY-1. Appl
602 Environ Microbiol **73:**7515-7521.
603 30. **Krivobok S, Kuony S, Meyer C, Louwagie M, Willison JC, Jouanneau Y.** 2003.
604 Identification of pyrene-induced proteins in *Mycobacterium* sp. 6PY1 : Evidence for
605 two ring-hydroxylating dioxygenases. J. Bacteriol. **185:**3828-3841.

606 31. **Parales RE, Lee K, Resnick SM, Jiang HY, Lessner DJ, Gibson DT.** 2000.
607 Substrate specificity of naphthalene dioxygenase: Effect of specific amino acids at the
608 active site of the enzyme. J. Bacteriol. **182:**1641-1649.
609 32. **Jones MD, Crandell DW, Singleton DR, Aitken MD.** 2011. Stable-isotope probing
610 of the polycyclic aromatic hydrocarbon-degrading bacterial guild in a contaminated
611 soil. Environ Microbiol **13:**2623-2632.
612 33. **Pinyakong O, Habe H, Omori T.** 2003. The unique aromatic catabolic genes in
613 sphingomonads degrading polycyclic aromatic hydrocarbons (PAHs). J. Gen. Appl.
614 Microbiol. **49:**1-19.
615 34. **Romine MF, Stillwell LC, Wong KK, Thurston SJ, Sisk EC, Sensen C,**
616 **Gaasterland T, Fredrickson JK, Saffer JD.** 1999. Complete sequence of a 184-
617 kilobase catabolic plasmid from *Sphingomonas aromaticivorans* F199. J Bacteriol
618 **181:**1585 - 1602.
619 35. **Pinyakong O, Habe H, Yoshida T, Nojiri H, Omori T.** 2003. Identification of three
620 novel salicylate 1-hydroxylases involved in the phenanthrene degradation of
621 *Sphingobium s*p. strain P2. Biochem Biophys Res Commun **301:**350 - 357.
622 36. **Kweon O, Kim SJ, Freeman JP, Song J, Baek S, Cerniglia CE.** 2010. Substrate
623 specificity and structural characteristics of the novel Rieske nonheme iron aromatic
624 ring-hydroxylating oxygenases NidAB and NidA3B3 from *Mycobacterium*
625 *vanbaalenii* PYR-1. mBio **1:**1-11**.**
626 37. **Yu CL, Liu W, Ferraro DJ, Brown EN, Parales JV, Ramaswamy S, Zylstra GJ,**
627 **Gibson DT, Parales RE.** 2007. Purification, characterization, and crystallization of
628 the components of a biphenyl dioxygenase system from *Sphingobium yanoikuyae* B1.
629 J. Ind. Microbiol. Biotechnol. **34:**311-324.

38. **Schuler L, Jouanneau Y, Chadhain SM, Meyer C, Pouli M, Zylstra GJ, Hols P, Agathos SN.** 2009. Characterization of a ring-hydroxylating dioxygenase from phenanthrene-degrading *Sphingomonas* sp. strain LH128 able to oxidize benz[a]anthracene. Appl Microbiol Biotechnol **83:**465-475.

39. **Furukawa K, Suenaga H, Goto M.** 2004. Biphenyl dioxygenases: Functional versatilities and directed evolution. J Bacteriol **186:**5189-5196.

40. **Pinyakong O, Habe H, Supaka N, Pinpanichkarn P, Juntongjin K, Yoshida T, Furihata K, Nojiri H, Yamane H, Omori T.** 2000. Identification of novel metabolites in the degradation of phenanthrene by *Sphingomonas* sp strain P2. Fems Microbiol. Lett. **191:**115-121.

41. **Seo JS, Keum YS, Hu Y, Lee SE, Li QX.** 2007. Degradation of phenanthrene by *Burkholderia* sp. C3: initial 1,2- and 3,4-dioxygenation and meta- and ortho-cleavage of naphthalene-1,2-diol. Biodegradation **18:**123-131.

42. **Moreno-Ruiz E, Hernaez MJ, Martinez-Perez O, Santero E.** 2003. Identification and functional characterization of *Sphingomonas macrogolitabida* strain TFA genes involved in the first two steps of the tetralin catabolic pathway. J Bacteriol **185:**2026-2030.

Table 1 : Bacterial strains and plasmids used in this study

Stain or plasmid	Description / genotype	Reference /source
Escherichia coli		
NEB 5-α	*fhuA2* Δ(*argF-lacZ)U169 phoA glnV44* Φ*80*Δ (*lacZ*)M15 *gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
BL21(DE3)	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Merckmillipore Novagen
Plasmids		
pJET1.2	Amp^R, cloning plasmid	Thermo Fisher scientific
pET15b	Amp^R, expression plasmid	Merckmillipore
pIZ1036	Km^R , broad range expression plasmid,	(42)
pIBA34	pIZ1036 carrying *phnA3* and *phnA4* from strain CHY-1	(18)
pJCA1	pJET1.2 carrying *pahAb*	This study
pJCA3	pJET1.2 carrying *pahAa*	This study
pJCA5	pJCA3 carrying *pahAb* from pJCA1 downstream of *pahAa*	This study
pJCA6/pJCA7	pJET1.2 carrying two different *pahAc5/pahAd5* amplicons	This study
pJCA8/pJCA9	pJET1.*2* carrying two different *pahAc4/pahAd4* amplicons	This study
pJCA10/pJCA11	pJET1.2 carrying two different *pahAc2/pahAd2* amplicons	This study
pJCA13/pJCA14	pJET1.2 carrying two different *pahAc8/pahAd8* amplicons	This study
pCAE1	pET15b carrying *pahAc2/pahAd2* from pJCA10	This study
pCAE4	pET15b carrying *pahAc8/pahAd8* from pJCA14	This study
pCAE5	pET15b carrying *pahAc4/pahAd4* from pJCA9	This study

pCAE7	pET15b carrying *pahAc5/pahAd5* from pJCA7		This study
pCAZ1	pIZ1036 carrying the *pahAa/pahAb* insert from pJCA5		This study

649

650 Table 2 : Oligonucleotides used for PCR amplification of RHD genes

Designation	5'-> 3' Sequence [a]	Targeted genes
C095-F	GAAGGAGATATATTATGGATAATAATTTCACCATC	*pahAa*
C095-R	C*GTCGAC*TTAATAAAAACGGTCAAAATGC AACTG	*pahAa*
C113-F	G*CTCGAG*GAGATATTGTATGAATACCCGTGTCAA	*pahAb*
C113-R	CTTATTCTACTTCTGCCTCAA	*pahAb*
C951-F	G*CATATG*ATAAATATAGATGATCTGATTGA	*pahAc2Ad2*
C341-R [b]	*CTCGAG*TTAAAATAAAGTGTTCATGTTGCTATC	*pahAc2Ad2*
C763-F	G*CATATG*ATGAAGCCAAGCGAGTTGATTGA	*pahAc4Ad4*
C569-F	G*CATATG*GTCGATGTAAACAGTCTG	*pahAc3Ad3*
C2271-R	GCCTACCCAATGGCTGATGCC	*pahAc3Ad3*
C451-F1	GAAGGAGATAT*CATATG*AATGAATGGCTGGAGGAG	*pahAc5Ad5*
C451-R	**G***TCTAGA*TCAGAAAAACATATTCAGATTTTTATC	*pahAc5Ad5*
C451-F2 [c]	GAAGGAGATAT*CATATG*AAAAA**C**aTtAACTATCAGGAAC	*pahAc5Ad5*
JCA7-R2 [c]	**GG***CTCGAG*ATCAGAAAAACATATTCAGA	*pahAc5Ad5*
C5241-F	G*CATATG*TTCGATATCAAGAATTTAATCAA	*pahAc8Ad8*
C427-R2	*CTCGAG*TTACAAGATAAACAACAAGTTTTTCCC	*pahAc8Ad8*

651

652 [a] Letters in italics indicate restriction sites for one of the following enzymes: NdeI, SalI,
653 XhoI or XbaI.

654 [b] This primer was also used for the amplification of *pahAc4Ad4* together with C763-F.

655 [c] Primers used for subcloning *pahAc5/Ad5* in pET15b.

656

Table 3 : Affiliation of 16S ribosomal sequences identified in the metagenomic data set obtained from labeled DNA

Proposed affiliation	Contig	Length (bp)	Best match accession	Query cover (%)	Identity (%)	Relevant OTU[a]
Acidobacteria						
unclassified	815	1275	FQ659841	100	96	OTU397
unclassified	9991	531	JN178653	67	97	NF
Alphaproteobacteria						
Bradyrhizobiaceae	6706	598	JN869027	86	94	NF
Betaproteobacteria						
Rhodocyclaceae	6757	600	FQ660427	100	99	OTU17
	8969	551	FQ660504	100	100	OTU101
Burkholderiales						
Comamonadaceae	10394	520	FQ660439	100	98	OTU6
	10648	519	FQ659005	100	99	OTU154
Gammaproteobacteria						
unclassified	1848	929	FQ660401	100	99	OTU2
Sinobacteriaceae	2378	848	FQ660299	100	99	OTU153
Xanthomonadaceae	3372	740	JN868994	87	93	NF
Xanthomonadales	8678	548	EF632898	67	97	
			FQ658754	38	96	OTU153

[a] Refers to OTUs previously associated to phenanthrene degraders as deduced from SIP experiments on the same study site (7). NF : no relevant OTU found

663 Table 4 – Properties of dioxygenase sequences cloned from metagenomic DNA

Genes[a]	Plasmids	Relevant contig	Mismatches vs contig	Product length	Percent identity[b]
pahAc2	pJCA10/11	341	178/1863	454	94.3
pahAd2				168	80.2
pahAc4	pJCA8/9	763 & 332	21/1852	454	87.7
pahAd4				168	77.5
pahAc5 (1)	pJCA6	451	235/1327	459	90.4
pahAd5 (1)				180	82
pahAc5 (2)	pJCA7	451	16/1327	453	89
pahAd5(2)				183	78.8
pahAc8	pJCA13/14	427	9/1653	449	92.3
pahAd8				176	86.4

664 [a] Each pair of RHD genes is represented by two different amplicons cloned in the
665 plasmids indicated in column 2. Amplicon sequences are closely similar except for the two
666 copies of *pahAc5/pahAd5*, which show 321 mismatches.
667 [b] Amino acid sequence identities with homologous gene products previously described by
668 Singleton et al. (12).
669

Table 5 : Dependence of recombinant dioxygenase activity on the co-expression of appropriate electron carriers

Oxygenase	Activity[a] with		
	PahAa/PahAb	PhnA4/PhnA3	No electron carrier
PahAc2/Ad2	2.27 ± 0.25	3.39 ± 0.49	$< 5\ 10^{-3}$
PahAc4/Ad4	0.118 ± 0.035	2.63 ± 0.59	0.0254 ± 0.0012
PahAc5/Ad5	0.720 ± 0.31	nd[b]	0.033 ± 0.0094
PahAc8/Ad8	2.44 ± 0.36	nd	0.054 ± 0.0072

[a]Activities are expressed as micromoles of dihydrodiol formed per hour per ml of culture normalized to an OD_{600} of 1.0. Naphthalene was used as substrate, except for PahAc5Ad5 and PahAc8Ad8, which were assayed with phenanthrene

[b]not determined

677 Table 6 : Activities of recombinant RHDs towards 2- to 4-ring PAHs
678

Substrate	**Product formed**	Dioxygenase activity [a]			
		PahAc2/Ad2	PahAc4/Ad4	PahAc5/Ad5	PahAc8/Ad8
Naphthalene	*cis*-1,2-dihydrodiol	1.34 ± 0.40	2.63 ± 0.59	2.65 ± 0.44	3.76 ± 0.18
Biphenyl	*cis*-2,3-dihydrodiol	-	-	-	3.75 ± 0.70
Phenanthrene	*cis*-3,4-dihydrodiol	0.173 ± 0.001	0.0228 ± 0.0081	0.113 ± 0.016	0.30 ± 0.075
	cis-1,2-dihydrodiol	0.806 ± 0.17	0.765 ± 0.28	0.238 ± 0.066	
Anthracene	*cis*-1,2-dihydrodiol	traces	traces	0.0381 ± 0.0058	0.529 ± 0.059
Fluorene[b]	dihydrodiol	-	-	-	$9.1 \pm 1.9\ 10^{-3}$
Pyrene	*cis*-4,5-dihydrodiol	-	-	$1.42 \pm 0.21\ 10^{-3}$	-

679 [a]Activities are expressed as micromoles of dihydrodiol formed per hour per ml of culture normalized to an OD_{600} of 1.0. (-): no detectable activity

680 [b] Fluorene oxidation by PahAc8/Ad8 also yielded monohydroxy- and dihydroxyfluorene, which were detected as trimethylsilylated derivates (see
681 text)

682 **Legends to figures**

683

684 **Figure 1 :** Maps of contigs 095 and 113 containing the RHD-specific electron carrier genes
685 *pahAa* and *pahAb*. The genes identified in the two contigs have counterparts in contig 05431
686 found in a previously described pyrene-degrading bacterial consortium (12). The putative
687 function of genes is depicted by different filling paterns as indicated

688

689 **Figure 2:** Overexpression of RHD components in recombinant *E. coli* strains as illustrated by
690 SDS-PAGE. Whole cell extracts were prepared from IPTG-induced culture normalized to a
691 bacterial density of 2.0 (OD_{600}). Samples (5 µl) were analyzed by slab gel electrophoresis
692 followed by Coomassie blue staining. The following RHDs were expressed in indicated
693 strains: lane 1: PahAc2/Ad2 in BL21(pCAE1)(pIBA34); lane 2: PahAc4/Ad4 in
694 BL21(pCAE5)(pIBA34); lane 3: PahAc5/Ad5 in BL21(pCAE7)(pCAZ1); lane 4:
695 PahAc8/Ad8 in BL21(pCAE4)(pCAZ1); lane 5: control strain BL21(pET15b)pCAZ1). Lane
696 6 shows a BL21(pJCA5) extract overexpressing PahAa ($M_r \approx 37{,}000$) and PahAa ($M_r \approx$
697 13,000). Lane 7 shows an extract of uninduced BL21(pCAE4)(pCAZ1). Scale on the right
698 indicates molecular mass markers in kilodaltons.

699

700 **Figure 3** : Phylogenetic tree showing the relationships between alpha subunit sequences of
701 selected PAH dioxygenases. RHD sequences studied in this work are presented in boldface
702 letters. Accession numbers to Genebank are indicated between brackets. Numbers at the
703 nodes indicate neighbor-joining bootstrap confidence. The sequence of the salicylate
704 hydroxylase from *Shingomonas* sp. CHY-1 (Phna1b) was used as an outgroup.

705

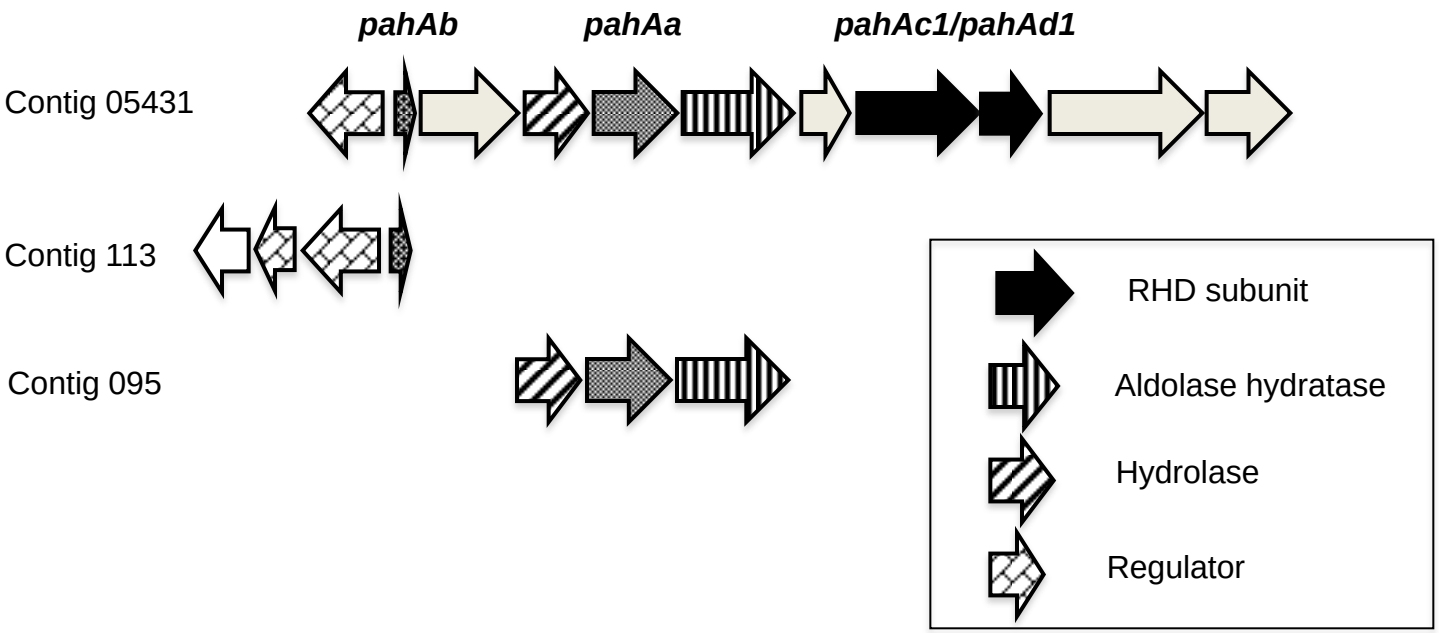

Figure 1, Chemerys et al.

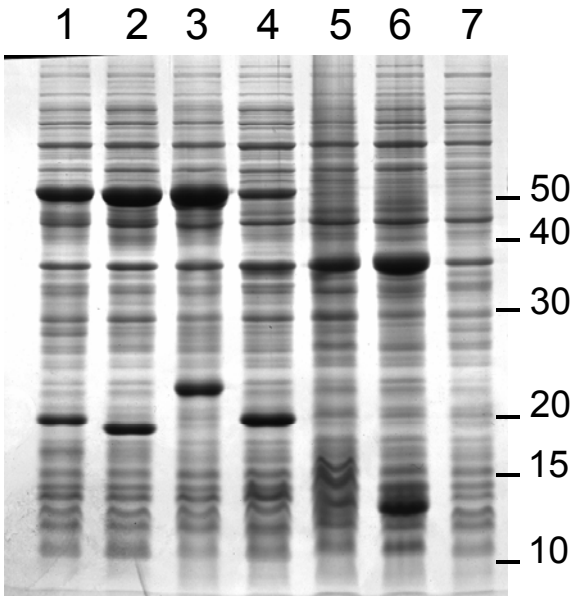

Figure 2: Chemerys et al.

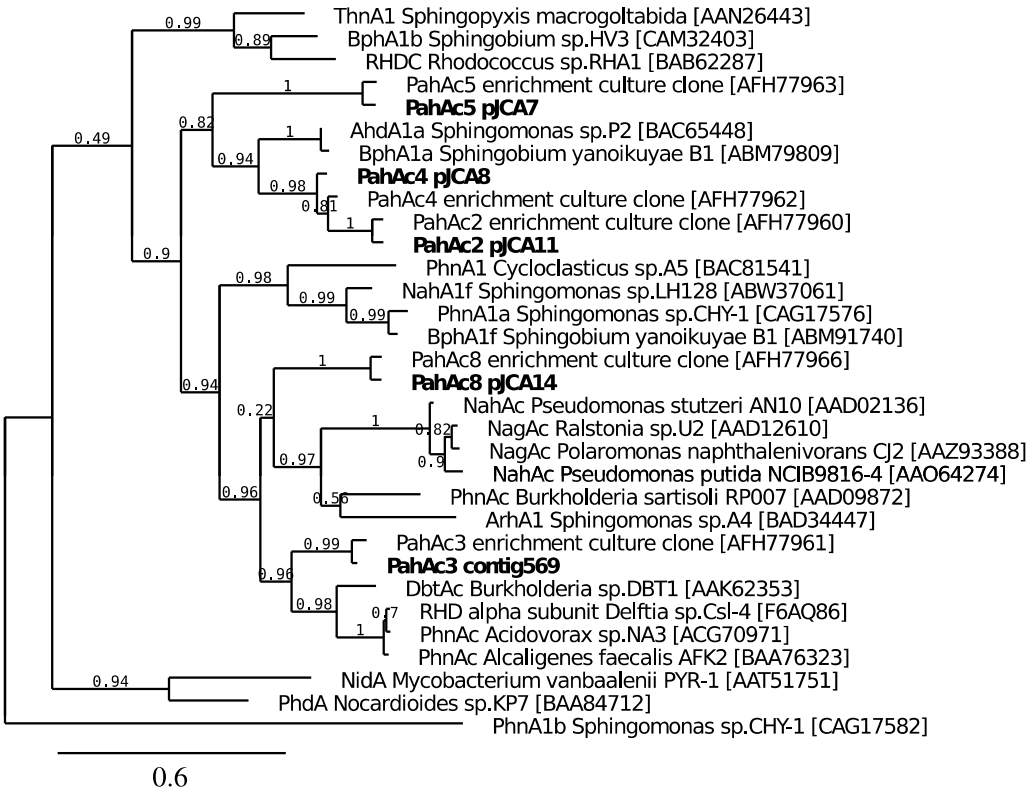

Figure 3 : Phylogenetic tree showing the relationships between alpha subunit sequences of selected PAH dioxygenases. RHD sequences studied in this work are presented in boldface letters. Accession numbers to Genebank are indicated between brackets. Numbers at the nodes indicate neighbor-joining bootstrap confidence. The sequence of the salicylate hydroxylase from *Shingomonas* sp. CHY-1 (Phna1b) was used as an outgroup.